

Indirect Measurement of Sexual Orientation: Comparison of the Implicit Relational Assessment Procedure, Viewing Time, and Choice Reaction Time Tasks

Jelena Rönspies,^{1,3} Alexander F. Schmidt,^{1,2} Anna Melnikova,¹ Rosina Krumova,¹ Asadeh Zolfagari,¹
and Rainer Banse¹

(this preprint has been accepted for publication in *Archives of Sexual Behavior*)

¹University of Bonn, Institute for Psychology, Social and Legal Psychology, Bonn, Germany

²University of Luxembourg, Institute for Health and Behaviour, INSIDE Research Unit, Walferdange, Luxembourg

³Correspondence concerning this article should be addressed to Jelena Rönspies, Department of Psychology, Social and Legal Psychology, University of Bonn, Kaiser-Karl-Ring 9, 53111 Bonn, Germany. Email: jeroe@uni-bonn.de.

ABSTRACT

The present study was conducted to validate an adaptation of the Implicit Relational Assessment Procedure (IRAP) as an indirect latency-based measure of sexual orientation. Furthermore, reliability and criterion validity of the IRAP were compared to two established indirect measures of sexual orientation: a Choice Reaction Time task (CRT) and a Viewing Time (VT) task. A sample of 87 heterosexual and 35 gay men completed all three indirect measures in an online study. The IRAP and the VT predicted sexual orientation nearly perfectly. Both measures also showed a considerable amount of convergent validity. Reliabilities (internal consistencies) reached satisfactory levels. In contrast, the CRT did not tap into sexual orientation in the present study. In sum, the VT measure performed best, with the IRAP showing only slightly lower reliability and criterion validity, whereas the CRT did not yield any evidence of reliability or criterion validity in the present research. The results were discussed in the light of specific task properties of the indirect latency-based measures (task-relevance vs. task-irrelevance).

KEY WORDS: Implicit Relational Assessment Procedure, Viewing Time, Choice Reaction Task, sexual orientation

INTRODUCTION

When a person is asked about personal attributes that affect intimate or socially sensitive domains, answers may not always be honest and distorted by attempts to present oneself in a way that is socially desired (Paulhus & Reid, 1991). Sexual orientation, described as a specific pattern of sexual interest in the opposite and/or same sex, may constitute such a sensitive domain, as sexual orientation statements indicating non-normative (i.e., non-heterosexual) interests may cause rejection or sanctioning in some social contexts (e.g., Tilcsik, 2011). Questionnaires and other self-report measures, therefore, have been criticized for their potentially decreased validity in assessing socially sensitive psychological constructs. Beyond the assessment topic's social sensitivity, another important limitation of self-report measures of sexual orientation is that the construct of sexual orientation may be rather multifaceted and dynamic than categorical (Klein, 1990) and thus more complex than the categorical labels used to describe sexual orientation (such as gay, lesbian, bisexual or heterosexual) suggest. This is corroborated by the various operationalizations of sexual orientation used in the literature, as for example, romantic/sexual attraction or arousal, sexual behavior, and sexual identity (Savin-Williams, 2006). Consequently, it has been shown that a considerable percentage of persons identifying themselves as heterosexual also occasionally experience sexual attraction to same-sex persons or engage in homosexual fantasies or contacts (e.g., Knight & Hope, 2012). Thus, explicit statements of sexual orientation—even though representing the sexual orientation label the person identifies with—may reflect only those aspects of sexual attraction, interest, fantasies, or behavior the individual is willing to be open about and that are introspectively accessible to a person or that are in accordance with the person's sexual identity.

Facing the drawbacks of self-report measures, psychological research has focused on the development of indirect measures of sexual interest as either based on physiological

reactions (i.e., penile plethysmography or phallometry) or, more recently, implicit social cognition. Although the former assessment approach has been established as a valid measure of male sexual interest (Harris & Rice, 1996), it is quite laborious and for most participants not very pleasant in terms of its application—a possible reason why it is not widely used in research with non-clinical/non-forensic populations. In contrast, the latter measurement approach most frequently relies on the assessment of response latencies in tasks that require fast reactions (e.g., rating, detecting or sorting) to stimuli that relate to the construct of interest. Besides its comparably easy application, the psychometric success of indirect latency-based measures has been attributed to the fact that these approaches benefit from being inherently less transparent than self-report measures and that indirect measures tap into automatic attitudes and behavioral dispositions that are not necessarily accessible to introspection.

In response to the specific problems of direct and physiological sexual interest measures (Kalmus & Beech, 2005), research interest has increasingly shifted to the indirect assessment of sexual interests using latency-based measures (e.g., Imhoff et al., 2010; Snowden, Wichter, & Gray, 2008; Wright & Adams, 1994). Thus, a considerable variety of conceptually different indirect measures tapping into (deviant) sexual interests has been developed (Snowden, Craig, & Gray, 2011; Larue et al., 2014; Schmidt, Banse, & Imhoff, *in press*). In the present study, three indirect latency-based measures of sexual orientation—including a newly developed task—will be compared, representing two conceptually different types of indirect measurement rationales: task-relevant and task-irrelevant indirect latency-based measurement of sexual orientation (Schmidt et al., *in press*).

Task-Relevant Measures

The term “task-relevant measure” (as opposed to “task-irrelevant” measure, see below) is used to label a subgroup of indirect latency-based measures in which participants

are required to explicitly react to sexually relevant features of target stimuli (Schmidt et al., in press). Target stimuli in task-relevant measures usually are categorized in terms of their sexual relevance or are to be classified into combinations with sexual attributes. Thus, participants are required to actively process sexual interest or relevance throughout task completion, rendering the task's assessment aim rather transparent to the individual. However, because indirect inferences about the construct of interest are drawn indirectly from behavior exhibited in response to stimuli, the task qualifies as an indirect measure.

The Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998) probably constitutes the best-researched example for task-relevant latency-based measures. The IAT is assumed to measure the relative strength of implicit associations between concepts and has been adapted to assess sexual orientation (e.g., Banse, Schmidt, & Clabour, 2010; Imhoff, Schmidt, Bernhardt, Dierksmeier, & Banse, 2011; Snowden et al., 2008; Snowden & Gray, 2013). The IAT's measurement rationale in measuring sexual orientation is based on the assumption that sexual interest results in the development of implicit associations between sexual attractiveness and the categories of preferred sexual partners. Though the IAT is one of the most applied indirect measures in implicit social cognition research, the measure's most discussed weakness is its relativity (De Houwer, 2002). As the IAT presumably assesses the relation between two association strengths, it does not provide a precise determination of an implicit association's nature or direction.

A task-relevant measure that is possibly able to overcome the conceptual weaknesses of the IAT (i.e., resulting in a relative instead of an absolute measure of concept relatedness) is the Implicit Relational Assessment Procedure (IRAP) (Barnes-Holmes et al., 2006). The IRAP is based on the relational frame theory (Barnes-Holmes, Hayes, Barnes-Holmes, & Roche, 2002), which postulates stimulus relations (as opposed to associations) to be central to human beliefs. In contrast to the IAT, the IRAP involves relational terms as it instructs the

participant to respond with reactions that correspond to the verbal relations underlying the participant's beliefs. As instructions alternate between different experimental blocks, participants are required to respond in a way that can be either consistent or inconsistent with their personal beliefs. For the present study, this meant that participants were asked to accept or deny simple statements such as "Woman-sexually attractive" or "Man-sexually unattractive". Similar to the IAT, the IRAP effect then leads to higher response latencies in inconsistent than in consistent trial blocks. The IRAP effect is hypothesized to result from an immediately initiated relative response that is elicited by the presented stimuli and can either match with the required response (i.e., decreased response latency) in the task or not (i.e., increased response latency). An important advantage of the measure is that—besides the overall difference score between consistent and inconsistent blocks—it allows to compute single difference scores for all four trial types, which are equivalent to the assessment of four single beliefs.

Since the first set of studies utilizing the IRAP (Barnes-Holmes et al., 2006), the measure has been shown to differentiate well between groups in known-group approaches (Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009) and has been validated to measure several psychological attributes, such as attitudes about meat (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010), self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009) or implicit beliefs of sexual offenders with child victims (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009). However, the IRAP has not been tested to assess sexual orientation. Thus, the present study involved an adapted version of the IRAP as a measure of sexual orientation.

In contrast to the relatively new IRAP, the by far oldest and most researched paradigm of task-relevant indirect latency-based measures of sexual interest is the Viewing Time (VT) task. Its rationale traces back to Rosenzweig (1942), who found that sexual

photographs were watched longer than non-sexual pictures. Typically, a sexual orientation VT task involves ratings of target persons differing in gender and therefore in regard to their sexual attractiveness (Imhoff et al., 2010). During task completion, response latencies are recorded unobtrusively. Higher mean response latencies in trials with targets of specific sex or sexual maturity are then interpreted as indicating the sexual interests of the individual. In spite of profound support for the validity of VT tasks in assessing sexual orientation (e.g., Bourke & Gormley, 2012; Ebsworth & Lalumière, 2012; Imhoff et al., 2010; Imhoff et al., 2011, Study 1; Israel & Strassberg, 2009; Lippa, 2012; Quinsey, Ketsetzis, Earls, & Karamanoukian; 1996; Zamansky, 1956), the underlying mechanisms have not been finally clarified. However, Imhoff et al. (2010) experimentally ruled out deliberate delay of response latencies and interfering attentional adhesion effects due to the sexually attractive nature of relevant target stimuli. Moreover, it could be shown that by far the greatest share of variance of VT effects was dependent on features of the task but not on characteristics of the stimuli (Imhoff, Schmidt, Weiß, Young, & Banse, 2012).

Task-Irrelevant Measures

In opposition to task-relevant measures, task-irrelevant indirect latency-based measures feature as primary task the detection of characteristics of target stimuli (e.g., semantic meaning, location, color) that are sexually irrelevant. Sexually relevant stimuli are then usually presented as distractors, simultaneously with the target stimuli (e.g., in the background of target stimuli). According to the underlying rationale of task-irrelevant measures, background stimuli are supposed to distract the individual's attention and interfere with the primary detection task if distractors are sexually relevant. Thus, the capture of attention due to the sexually relevant distractor stimuli results in an increase of response latencies to the respective target stimuli (Schmidt et al., in press).

The Choice Reaction Time (CRT) (Wright & Adams, 1994; 1999) constitutes a task-irrelevant latency-based measure of sexual orientation, as it relies on the assumption that the presentation of sexually attractive stimuli interferes with the completion of a primary reaction task due to attentional resources that are captured by the sexually attractive nature of the distractors. In typical CRT tasks, participants are instructed to locate a target stimulus that is superimposed on pictures of sexually preferred or non-preferred persons by pressing a button corresponding to the targets' position in the picture. Participants exhibit longer response latencies in CRT tasks when the pictures show sexually attractive individuals. CRT tasks have been shown to measure sexual orientation in a smaller number of studies (Santtila et al., 2009; Wright & Adams, 1994; 1999). However, Santtila and colleagues reported methodological difficulties as they found the CRT to be a valid measure of sexual preference only when a subset of measures was taken into account. Out of three blocks, each consisting of 160 trials, a CRT effect could only be found for the first block, which was discussed as a possible task habituation effect. Thus, the current body of literature available on the measurement of sexual orientation using CRT tasks is not very consistent and partially depends on methodological alterations of the measure. To provide robust evidence for the validity of the CRT in this domain, further replications are needed.

This inconsistency of empirical evidence is not limited to the CRT. In general, task-irrelevant latency based measures have been discussed to be problematic in assessing sexual interest. In spite of several successful applications in clinical contexts not related to sexuality with negative, threatening stimuli serving as source for attention capture (Cisler, Bacon, & Williams, 2009), the rationale underlying task-irrelevant indirect measures can be used to derive theoretically opposing expectations when applied to sexual interest indicators (as, for example, argued by Prause, Janssen, & Hettrick, 2008). Typical attention-based measures cannot be used to differentiate between initial attention capture and subsequent difficulties in

disengagement from the salient stimulus (Fox, Derakshan, & Standage, 2011). Moreover, the reliability of task-irrelevant latency-based measures has turned out to be problematic (Cisler et al., 2009) and data concerning the reliability of task-irrelevant measures of individual differences in sexual interest are not available so far.

From a methodological perspective, if one seeks to predict behavior from a behavioral measure, predictive validity should be best when the predictive task is structurally similar to the behavior to be predicted (Perugini, Richetin, & Zogmaister, 2010). Therefore, task-relevant indirect measures have the advantage that the predictive measure is based upon the explicit processing of sexual relevance of the stimuli in the primary task whereas in task-irrelevant measures the primary task (e.g., color classification) is not directly related to the criterion behavior. Hence, the task-relevant measures of sexual relevance can be expected to be structurally more equivalent to the classification of sexual orientation as compared to task-irrelevant tasks. This account has received preliminary support from indirect latency-based measures that are based on response-interference processes (Gawronski, Deutsch, LeBel, & Peters, 2008; Suchotzki, Verschuere, Crombez, & DeHouwer, 2013).

CURRENT STUDY

The present study focused on two research aims: First, concerning the IRAP, our main objective was to validate the IRAP for assessing sexual orientation as it has not been applied in this domain before. Second, by simultaneously assessing two further conceptually different indirect latency-based measures of sexual orientation that were based either on task-relevant or task-irrelevant measurement rationales, we aimed to compare these measures in terms of psychometric properties (i.e., reliability, criterion validity). We chose variants of the well-validated VT task and a newly developed version of the IRAP as task-relevant latency-based measures and an adaptation of the CRT as a task-irrelevant latency-based measure because all three indirect latency-based measures were based on non-relative comparisons of

absolute sexual interest indicators (as opposed to the well-validated IAT measures of sexual interest). Utilizing a known-group approach within a sample of self-identified homo- and heterosexual men, we hypothesized that the three measures would be valid predictors of sexual orientation. However, we expected differential patterns of psychometric properties. More precisely, the task-relevant measures (VT, IRAP) were hypothesized to outperform the task-irrelevant paradigm (CRT) in terms of criterion validity and reliability as task-relevant measures of sexual interest tend to be psychometrically superior (Schmidt et al., in press).

METHOD

Participants

The sample consisted of 87 heterosexual and 35 gay men recruited via online forums, social media platforms, and posters to participate in an online study on sexual attractiveness. Sexual orientation was assessed using the Kinsey Scale (Kinsey, Pomeroy, & Martin, 1948). Only participants who rated themselves as either unambiguously gay (Kinsey Scale 5 and 6) or heterosexual (Kinsey Scale 0 and 1) were included. Participant age ranged from 18 to 53 years ($M = 27.58$, $SD = 7.73$); gay men ($M = 30.31$, $SD = 8.07$) were significantly older than heterosexual men ($M = 26.47$; $SD = 7.34$), $t(119) = -2.54$, $p < .05$ ¹. Prior to the experiment, all participants were informed that pictures showing nudity and measures of sexual interest would be part of the study. Participants were further informed, that they could terminate their participation at any time during the experiment without any personal disadvantage and that data would be recorded anonymously.

Measures

Implicit Relational Assessment Procedure (IRAP)

¹ Three participants did not provide any information regarding their age.

The original IRAP task (Barnes-Holmes et al., 2006) was modified to assess sexual orientation. In a total of six experimental blocks, participants were instructed to either agree or disagree (by pressing the *X* key for *right* and the *M* key for *wrong* to differing combinations of word stimuli that were presented on screen in each trial. Target stimuli were 12 German words associated to *sexually attractive* (e.g., *erotic*, *seductive*, *hot*) or *sexually unattractive* (e.g., *non-erotic*, *dull*, *boring*). Target stimuli were simultaneously presented with gender category stimuli (i.e., *woman*, *man*) on screen, resulting in four trial combinations (*man-sexually attractive*, *woman-sexually attractive*, *man-sexually unattractive*, *woman-sexually unattractive*).

Instructions prior to each experimental block alternately requested participants to respond with *right* and *wrong* to different trial combinations. In Blocks 1, 3, and 5, participants were instructed to respond to *woman-sexually attractive* and *man-sexually unattractive* with *right*, and to respond with *wrong* when combinations of *man-sexually attractive* or *woman-sexually unattractive* were presented. This response pattern was congruent with a male heterosexual orientation. In Blocks 2, 4, and 6, instructions were reversed and participants were requested to respond in a gay-congruent way (*man-sexually attractive* and *woman-sexually unattractive* as *right*, *woman-sexually attractive* and *man-sexually unattractive* as *wrong*). Participants were instructed to respond as quickly and accurate as possible. All experimental blocks consisted of 24 trials per block with 6 trials per trial type. Trials were followed by an interval of 400 ms before the next trial was presented. Incorrect responses were indicated by a red *X* presented in the center of the screen and the next trial was not started until participants had corrected their response by pressing the correct answer key. Response latencies were recorded and error trials as well as latencies of more than 10,000 ms were treated as missing.

The *d*-IRAP scores were calculated as described by Barnes-Holmes et al. (2009). To this end, average latencies for the four trial types were estimated separately for each trial block. Mean latencies of heterosexual-congruent trial blocks were then subtracted from mean latencies of gay-congruent blocks and standardized by the pooled *SD*, resulting in 12 single difference scores that represented the four trial types for each pair of gay- and heterosexual-congruent blocks. Averaging the obtained difference scores separately for trial type resulted in four final *d*-IRAP scores. A fifth final overall *d*-IRAP score was estimated by jointly averaging all 12 difference scores. Thus, positive *d*-IRAP scores indicated a heterosexual-congruent pattern of response latencies, whereas negative *d*-IRAP scores indicated a response bias congruent with gay sexual interest. Moreover, larger *d*-IRAP scores (positive or negative) indicated greater differences between latencies of heterosexual- and gay-congruent trial blocks.

Viewing Time (VT)

As the VT task, a speeded version of the standard paradigm with restricted response windows was utilized (Imhoff et al., 2010, Study 3). Participants were instructed to complete a binary decision task by indicating whether a displayed target person was either a *potential sexual partner* or *not a potential sexual partner* from their subjective perspective. Participants were instructed to respond as quickly as possible within 1,000 ms.

Target stimuli were 80 pictures taken from the Not Real People (NRP) picture set (Pacific Psychological Assessment Corporation, 2004). The NRP features pictures of computer-generated persons in bathing clothes, who differ in sex and status of sexual maturation according to Tanner (1978). Pictures of persons belonging to Tanner categories 1, 2, and 3 depicted prepubescent boys and girls and Tanner categories 4 and 5 featured post-pubescent adolescents and adults. The set of stimuli consisted of 40 male and 40 female persons with each of the subsets containing eight pictures per Tanner category. Prepubescent

target stimuli functioned as distractors as these were not sexually relevant in a non-clinical community sample.

Target pictures were presented while reminders of the two response categories were displayed on both sides of the stimulus. To indicate that the presented target person would constitute a potential sexual partner, participants were instructed to press the *X* key on their keyboards, the opposite response was given by pressing the *M* key. As the response time of 1,000 ms expired, a “*too slow!*” feedback was displayed in the center of the target stimulus. Participants then had to respond before the next trial started. Response latencies were recorded unobtrusively and averaged across responses that were given < 1,000 ms for all stimulus categories. An overall standardized mean difference VT score was obtained by computing the difference between mean response latencies for post-pubescent male and female pictures (Tanner categories 4 and 5), divided by the pooled *SD* of the respective trials (*d*-measure).

Choice Reaction Time (CRT)

The stimuli in the CRT were 54 pictures taken from the Virtual People Set (Dombert et al., 2013) including 18 pictures of nude post-pubescent male and female persons (Tanner categories 4 and 5) and 36 pictures of nude prepubescent persons (Tanner categories 1, 2, and 3), which served as distracting stimuli. Six practice trials were presented before the experimental trials started. The main task consisted of 216 trials, in which pictures were presented with an orange dot displayed randomly in one of the four corners. Prior to each trial, a fixation cross was presented in the center of the screen for 500 ms. Participants were instructed to focus on the fixation cross and then locate the dot as quickly as possible by pressing the correct key on a keyboard (*R*, *U*, *V*, and *N* keys; assignment of the keys corresponded to the position of the four corners). Having responded after 1,000 ms, participants received a feedback on the screen indicating their response was too slow. In this

case, participants had to correct their response and the trial was counted as an error trial. Incorrect responses were followed by an error feedback and the next trial started without the requirement of correction. Response latencies were calculated exclusively from correct trials. Latencies for all stimulus categories were averaged and an overall standardized mean difference CRT score was computed by subtracting the mean latency of post-pubescent male stimuli trials from the mean latency of post-pubescent female stimuli trials, divided by the pooled *SD* (*d*-measure).

Procedure

Participants were recruited via postings in internet forums and social media networks (e.g., Men's Health, Facebook) and did not receive any compensation for partaking in the study except that they were provided the possibility to win one out of two internet shopping vouchers (each 20 € approximately 26 US \$). Potential participants were informed that the study was about sexual attractiveness. All measures were assessed online. At the beginning of the survey, informed consent was obtained and participants were asked about their age, sex, and sexual orientation. Subsequently, participants completed the IRAP, followed by the VT task and the CRT. In order to maximize individual differences (i.e., sexual orientation effects), stimuli were presented in the same random order for all participants, across all measures. After the assessment of the indirect latency-based measures was completed, participants were debriefed and thanked and could leave an email-address for the shopping voucher raffle.

RESULTS

Implicit Relational Assessment Procedure

Data from four heterosexual participants were eliminated due to occurrence of low latencies (< 300ms) in at least 10% of trials (Barnes-Holmes et al., 2009). Reliability of the IRAP was calculated by computing three difference scores for the three pairs of gay-

consistent and heterosexual-consistent trial blocks (Blocks 1 and 2, 3 and 4, and 5 and 6) and subsequently computing the internal consistency of the difference scores. Internal consistency reached a satisfactory level ($\alpha = .78$, Table 1).

As shown in Table 1, mean overall IRAP d -scores were positive for heterosexual men and negative for homosexual men. Thus, gay men showed lower reaction times for trial blocks that were consistent with sexual attraction to men while heterosexual men responded more quickly in trial blocks consistent with a sexual attraction to women. This latency difference was significant and corresponded to a large effect ($d = -1.58$).

Figure 1 depicts the average response latencies of both heterosexual and gay men across all four trial types. In a series of four t -tests, heterosexual and gay men were compared regarding their scores on each trial type. Results revealed significant group effects for all trial types (all $ps < .001$) with strong effect sizes between $d = 0.98$ and $d = 1.37$. A 2 (Sexual Orientation) \times 4 (Trial Type) mixed-model ANOVA with Sexual Orientation as a between-subjects factor and Trial Type as a within-subjects factor was conducted.² As the critical test for the validity of the IRAP, the interaction between Sexual Orientation and Trial Type yielded a marginally significant effect, $F(3, 112) = 2.5$ $p = .06$, $\eta^2 = .02$, indicating that both gay and heterosexual men showed lower response latencies when reacting to trial combinations that represented beliefs consistent with their sexual orientation (*man-sexually attractive* and *woman-sexually unattractive* for gay men and *woman-sexually attractive* and *man-sexually unattractive* for heterosexual men).

² As is the case for all latency-based measures, we had to hypothesize that reaction times may be affected by participants' age. We included age as a covariate in all our analyses in order to control for general age-reaction speed effects. For reasons of parsimony, results of the covariate analyses are only reported when significant. Unless otherwise reported, the covariate did not have a significant effect. However, in this case including the covariate—although non-significant—yielded a significant Sexual Orientation \times Trial Type interaction effect in the respective ANOVA reported above, $F(3, 112) = 3.09$ $p < .05$, $\eta^2 = .03$.

In order to test the IRAP's potential to correctly predict sexual orientation for gay and heterosexual men, a Receiver-Operator Characteristic (ROC) analysis was conducted. The IRAP *d*-score reached an *AUC* index (Area Under the Curve) of .90 ($p < .001$, Table 1), indicating that the IRAP differentiated well between heterosexual and gay men (with an *AUC* of 0.50 indicating a random hit rate and an *AUC* of 1.00 indicating a perfect hit rate).

Viewing Time

There were no indications of extreme responses from any participant that had to be removed from the analyses. Because prepubescent stimuli served as distractor tasks, reliability of the VT task was calculated from trials with post-pubescent stimuli exclusively (Tanner 4 and Tanner 5). Two standardized mean difference scores, each for one half of post-pubescent male and female stimuli, were computed by subtracting the average response latencies for male stimuli from the latencies for female stimuli, divided by the pooled *SD*.

The reliability of the two difference scores was good ($\alpha = .90$, Table 1). As shown in Table 1, gay men, on average, produced longer response latencies for male compared to female target stimuli, while response latencies of heterosexual men were higher in trials with female targets (indicated by negative overall VT *d*-scores for gay men and positive scores for heterosexual men). Again, this response pattern was significant and resulted in an even larger sexual orientation effect ($d = -3.59$, Table 1).

Critically, a 2 (Participant Sexual Orientation) x 2 (Target Sex) x 2 (Target Sexual Maturity) mixed-model ANOVA revealed that the crucial three-way interaction effect of Target Sexual Maturity, Target Sex, and Sexual Orientation was significant, $F(1, 115) = 46.57$, $p < .001$, $\eta^2 = .28$, indicating that participants showed longer latencies for targets that were potential sexual partners according to their sexual orientation and that this effect was stronger for sexually mature targets. The *AUC* of the VT task for the prediction of sexual orientation reached a near-perfect hit rate of .99 ($p < .001$, Table 1).

Choice Reaction Time

Three participants had to be excluded from the analysis because of extreme overall CRT scores ($> 3 SD$). Reliability of the CRT was analyzed by splitting the latencies for trials with sexually mature (Tanner categories 4 and 5) targets in two halves and for both halves subtracting the latencies for male from those for female targets. It was unacceptably low ($\alpha = .23$). Also, heterosexual and gay men did not differ significantly in their overall CRT scores (Table 1).

A 2 (Participant Sexual Orientation) x 2 (Target Sex) x 3 (Target Sexual Maturity) ANCOVA (with Participants Age as a covariate) resulted in a strong main effect of the covariate, $F(1, 115) = 32.61, p < .001, \eta^2 = .22$, and a smaller main effect of sexual orientation, $F(1, 115) = 4.60, p < .05, \eta^2 = .04$, indicating that participants reacted more slowly when they were older and that heterosexual men showed significantly higher response latencies than gay men. This means participants' CRT response latencies were independent from target stimulus characteristics (gender, sexual maturity level).³

Intercorrelations

Table 2 shows the correlations between the three measures utilized in the study, as well as participants' age and sexual orientation. The IRAP and VT were significantly and substantially correlated with participants' sexual orientation whereas no significant association with the self-report measure of sexual orientation was found for the CRT. Moreover, the CRT was not significantly correlated to the other two measures, whereas the IRAP and VT were significantly associated with each other at a substantial level and in the

³ In order to control for possible habituation or practice effects (as discussed by Santtila et al., 2009) due to the high number of trials in the present CRT, we also conducted all analyses separately for the first 108 and second 108 trials. The pattern found in the analysis of all trials still held for both halves of trials, with significant main effects of Age and Sexual Orientation, but no significant two-way or three-way interactions were obtained. Consequently, the *AUC* indices of both scores were not significant (*AUC* = .49 for the first 108 trials, and *AUC* = .56 for the second 108 trials).

expected positive direction. This pattern corroborated convergent validity for the IRAP and VT, but not the CRT. The pattern still held when controlling for the age-sexual orientation confound (Table 2, below the main diagonal), indicating that participants' performance on the latency based difference scores was largely unaffected by age effects (due to the ipsatization by calculating individual difference measures between sexual orientation-consistent and inconsistent trial types that inherently control for possible age confounds).

DISCUSSION

Our central objectives were to determine whether the task-relevant IRAP and the task-irrelevant CRT are valid measures to assess sexual orientation and to compare these paradigms against the well-validated VT in terms of their psychometric properties. Results revealed substantial evidence for the criterion validity of the VT and IRAP—as task-relevant measures—while the task-irrelevant CRT failed to show any significant association with sexual orientation.

As the best established indirect latency-based measure of sexual interest, the VT task showed high criterion validity differentiating nearly perfectly between heterosexual and gay men. We also found strong evidence for the convergent validity of the VT. Moreover, this is a first conceptual replication of the specific VT variant with restricted response windows introduced by Imhoff et al. (2010, Study 3) that was validated with heterosexual male and female participants exclusively in the original study.

Furthermore, the present research was the first application of the IRAP as a measure of sexual orientation. The analysis of IRAP data revealed that participants' response latencies were shorter when response instructions were consistent with the participants' sexual orientation. Thus, participants significantly differed in their IRAP *d*-score depending on their sexual orientation. However, there are two notable exceptions of the hypothesized response pattern: First, gay men's difference scores for the *woman-sexually unattractive* trials of the

IRAP were not significantly different from zero. It might be conceivable that some gay men may consider female sexiness as something positive and desirable, which may have resulted in less clear-cut latency differences in items associated to female attractiveness. As a second exception, heterosexual men's difference scores for "man-sexually attractive" trials were below zero. Thus, heterosexual men showed slightly lower latencies in homosexual-congruent blocks. A possible explanation might be that in contrast to gay sexuality, heterosexual sexuality involves both men and women. Thus, male attractiveness may be a relevant concept also for heterosexual men when processing sexual attractiveness.

Strikingly, the IRAP's accuracy in predicting group membership was only slightly lower than the perfect VT's and evidence of convergent validity could be obtained in terms of substantial correlations with explicitly assessed sexual orientation and the VT task as well. Overall, the IRAP was a reliable and valid measure of sexual interest and thus should be considered as an additional promising indirect latency-based measure of sexual orientation.

In contrast, the CRT task did not turn out as sufficiently reliable nor were sexual orientation-specific effects on reaction times found. It is puzzling that the CRT results showed a significant main effect of group, while no two-way or three-way interactions with any target characteristics reached significance. Hence, our CRT results were clearly in opposition to the results from Wright and Adams (1994; 1999) and Santtilla et al. (2009) who found evidence for the validity of a CRT task in differentiating between heterosexual and gay participants. A possible explanation of the differing results may be provided by methodological differences between our research and the other studies. Both Wright and Adams (1994) and Santtilla and colleagues (2009) added a third, neutral stimulus category in the CRT, which we substituted by using pictures of prepubescent persons as distracting stimuli. However, in our analyses, we did not find any significant effects of Target Sexual Maturity. Thus, the use of stimuli varying in sexual maturity may not account for the differing

results. Another methodological point to be discussed is that Santtilla and colleagues found the CRT (including a simultaneously presented priming task) to be differentiating between groups only, when exclusively the first set of trials was taken into account. The CRT conducted here comprised nearly twice as many trials as the CRT conducted by Wright and Adams (1994, 1999) and the first CRT block in the study by Santtilla et al. (2009). Nevertheless, we analyzed both halves of the present CRT trials separately in order to control for possible habituation, practice, or loss of test motivation effects. Taking into account only the first 108 trials, no significant interactions were revealed and the pattern of effects found when all 216 trials were included still held for both single halves of CRT trials. Future research thus should address the influences of methodological features on the CRT-task's validity in assessing sexual orientation. Notably, our test of the sexual orientation CRT relies on the largest sample published so far with especially twice as many gay men as in the abovementioned CRT studies. This increase in statistical power needs to be taken into account in the interpretation of our non-significant results.

It is a limitation to the present finding that our sample was limited to self-identified gay and heterosexual male participants exclusively. This was done to maximize individual differences in a known-group approach as female sexual interests are generally less specifically confined to one sex (e.g., Chivers, Rieger, Latty, & Bailey, 2004). However, future research should include female participants as a further test of the IRAP's criterion validity and replicate the finding of greater female non-specificity of female sexual orientation that has been reported for indirect latency-based measures as well (e.g., Imhoff et al., 2010; 2011, Study 3). Unfortunately, we were not able to strictly test the differential criterion validities of task-relevant vs. task-irrelevant indirect measures of sexual orientation as the CRT did not predict sexual orientation at all. However, our hypothesis of superior psychometric quality of task-relevant measures was supported by the substantially higher

reliability coefficients of the task-relevant measures (VT, IRAP) The present research thus adds evidence to the notion that psychometric properties of latency-based measures of sexual interest seem to benefit from an explicit processing of the sexual relevance of target stimulus properties (task-relevant measurement rationale). This is further corroborated by the finding that CRT results are less robust across studies as compared to task-relevant indirect sexual orientation/sexual interest measures such as VT and IAT (Schmidt et al., in press).

In summary, from the present results, the IRAP appears as a new promising indirect measure to assess sexual orientation as it provides an assessment of relational associations between concepts and allows to draw inferences about the underlying strength of associations (as opposed to purely relative measures of associations such as the IAT). Although the VT task performed slightly better than the IRAP, there are substantial conceptual differences between the measures, so that the IRAP as a measure of “beliefs” (Barnes-Holmes et al., 2006) could provide a useful attitudinal adjunct to the already validated and more established indirect measures of sexual orientation such as VT and IAT.

REFERENCES

- Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior*, 37, 319–335. doi:10.1177/0093854809357598
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, 32, 169–177.
- Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *Psychological Record*, 60, 287–306.
- Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test (IAT): Measuring attitudes towards Dublin and country life in Ireland. *Psychological Record*, 58, 389–406.
- Barnes-Holmes, Y., Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2002). Relational frame theory: A post-skinnerian account of human language and cognition. *Advances in Child Development and Behavior*, 28, 101–138. doi:10.1016/S0065-2407(02)80063-5
- Bourke, A. B., & Gormley, M. J. (2012). Comparing a pictorial Stroop task to viewing time measures of sexual interest. *Sexual Abuse: A Journal of Research and Treatment*, 24, 479-500. doi: 10.1177/1079063212438922

- Chivers, M. L., Rieger, G., Latty, E., & Bailey, J. M. (2004). A sex difference in the specificity of sexual arousal. *Psychological Science*, *15*, 736–744.
doi:10.1111/j.0956-7976.2004.00750.x
- Cisler, J. M., Bacon, A. K., & Williams, N. L. (2009). Phenomenological characteristics of attentional biases towards threat: A critical review. *Cognitive Therapy and Research*, *33*, 221–234. doi:10.1007/s10608-007-9161-y
- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment*, *21*, 57–75. doi:10.1177/1079063208326928
- De Houwer, J. D. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry*, *33*, 115–133. doi:10.1016/S0005-7916(02)00024-1
- Dombert, B., Mokros, A., Bruckner, E., Schlegl, V., Antfolk, J., Backstrom, A., Zappalà, A., . . . Santtila, P. (2013). The Virtual People Set: Developing computer-generated stimuli for the assessment of pedophilic sexual interest. *Sexual Abuse: A Journal of Research and Treatment*. Advance online publication.
doi:10.1177/1079063212469062
- Dubé, E. M. (2000). The role of sexual behavior in the identification process of gay and bisexual males. *Journal of Sex Research*, *37*, 123–132.
doi:10.1080/00224490009552029
- Ebsworth, M., & Lalumière, M. L. (2012). Viewing time as a measure of bisexual interest. *Archives of Sexual Behavior*, *41*, 161–172. doi: 10.1007/s10508-012-9923-9

- Fox, E., Derakshan, N., & Standage, H. (2011). The assessment of human attention. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 15–47). New York: Wiley.
- Gawronski, B., Deutsch, R., LeBel, E. P., Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures—some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, 24, 218-225. doi: 10.1027/1015-5759.24.4.218
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi:10.1037/h0087889
- Hammack, P. L. (2005). The life course development of human sexual orientation: An integrative paradigm. *Human Development*, 48, 267–290. doi:10.1159/000086872
- Harris, G. T., & Rice, M. E. (1996). The science in phallometric measurement of male sexual interest. *Current Directions in Psychological Science*, 5, 156-160. doi: 10.1111/1467-8721.ep11512355
- Imhoff, R., Schmidt, A. F., Bernhardt, J., Dierksmeier, A., & Banse, R. (2011). An inkblot for sexual preference: A semantic variant of the Affect Misattribution Procedure. *Cognition & Emotion*, 25, 676-690. doi:10.1080/02699931.2010.508260
- Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior*, 39, 1275–1288. doi:10.1007/s10508-009-9595-2
- Imhoff, R., Schmidt, A. F., Weiß, S., Young, A. W., & Banse, R. (2012). Vicarious viewing time: Prolonged response latencies for sexually attractive targets as a function of

- task- or stimulus-specific processing. *Archives of Sexual Behavior*, 41, 1389–1401.
doi:10.1007/s10508-011-9879-1
- Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior*, 38, 551–558.
doi:10.1007/s10508-007-9246-4
- Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior*, 10, 193-218. doi:10.1016/j.avb.2003.12.002
- Kinsey, A.C., Pomeroy, W.B., & Martin, C.E. (1948). *Sexual behavior in the human male*. Philadelphia, PA: Saunders.
- Klein, F. (1990). The need to view sexual orientation as a multivariable dynamic process: A theoretical perspective. In D.P. McWhirter, S. A. Sanders, & J. M. Reinish (Eds.), *Homosexuality/heterosexuality: Concepts of sexual orientation* (pp. 277–282). New York: Oxford University Press.
- Knight, L. F. M., & Hope, D. A. (2012). Correlates of same-sex attractions and behaviors among self-identified heterosexual university students. *Archives of Sexual Behavior*, 41, 1199–1208. doi:10.1007/s10508-012-9927-5
- Larue, D., Schmidt, A. F., Imhoff, R., Eggers, K., Schönbrodt, F. D., & Banse, R. (2014). Validation of direct and indirect measures of preference for sexualized violence. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000016
- Lippa, R. A. (2012). Effects of sex and sexual orientation on self-reported attraction and viewing times to images of men and women: Testing for category specificity. *Archives of Sexual Behavior*, 41, 149-160. doi: 10.1007/s10508-011-9898-y
- Pacific Psychological Assessment Corporation. (2004). *The Not-Real People Stimulus Set for assessment of sexual interest*. Victoria, BC, Canada: PPAC.

- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307–317. doi: 10.1037/0022-3514.60.2.307
- Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of social cognition: Measurement, theory, and applications* (p. 255-277). New York: Guilford.
- Prause, N., Janssen, E., & Hetrick, W. P. (2008). Attention and emotional responses to sexual stimuli and their relationship to sexual desire. *Archives of Sexual Behavior*, 37, 934–949. doi:10.1007/s10508-007-9236-6
- Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology*, 17, 341–354. doi:10.1016/S0162-3095(96)00060-X
- Rosenzweig, S. (1942). The Photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine*, 4, 150–158.
- Santtila, P., Mokros, A., Viljanen, K., Koivisto, M., Sandnabba, N. K., Zappalà, A., & Osterheider, M. (2009). Assessment of sexual interest using a Choice Reaction Time task and priming: A feasibility study. *Legal and Criminological Psychology*, 14, 65–82. doi:10.1348/135532507X267040
- Savin-Williams, R. C. (2006). Who's gay? Does it matter? *Current Directions in Psychological Science*, 15, 40-44. doi: 10.1111/j.0963-7214.2006.00403.x
- Schmidt, A. F., Banse, R., & Imhoff, R. (in press). Indirect measures in forensic contexts. In van de Vijver, F. J. R. & T. Ortner (Eds.), *Behavior based assessment in personality, social and applied psychology*. Göttingen, Germany: Hogrefe.

- Snowden, R. J., Craig, R. L., & Gray, N. S. (2011). Indirect behavioral measures of cognition among sexual offenders. *Journal of Sex Research*, 48, 192–217.
doi:10.1080/00224499.2011.557750
- Snowden, R. J., & Gray, N. S. (2013). Implicit sexual associations in heterosexual and homosexual women and men. *Archives of Sexual Behavior*, 42, 475–485.
doi:10.1007/s10508-012-9920-z
- Snowden, R. J., Wichter, J., & Gray, N. S. (2008). Implicit and explicit measurements of sexual preference in gay and heterosexual men: A comparison of priming techniques and the Implicit Association Task. *Archives of Sexual Behavior*, 37, 558–565.
doi:10.1007/s10508-006-9138-z
- Suchotzki, K., Verschuere, B., Crombez, G., & De Houwer, J. (2013). Reaction time measures in deception research: Comparing the effects of irrelevant and relevant stimulus-response compatibility. *Acta Psychologica*, 144, 224–231. doi:
10.1016/j.actpsy.2013.06.014
- Tanner, J. (1978). *Foetus into man: Physical growth from conception to maturity*. Cambridge, MA: Harvard University Press.
- Tilcsik, A. (2011). Pride and prejudice: Employment discrimination against openly gay men in the United States. *American Journal of Sociology*, 117, 586–626.
doi:10.1086/661653
- Vahey, N., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, 59, 371–388.
- Wright, L. W., & Adams, H. E. (1994). Assessment of sexual preference using a Choice Reaction Time task. *Journal of Psychopathology and Behavioral Assessment*, 16, 221–231. doi:10.1007/BF02229209

Wright, L. W., & Adams, H. E. (1999). The effects of stimuli that vary in erotic content on cognitive processes. *Journal of Sex Research*, 36, 145–151. Doi:

10.1080/00224499909551979

Zamansky, H. S. (1956). A technique for measuring homosexual tendencies. *Journal of Personality*, 24, 436-448. doi: 10.1111/j.1467-6494.1956.tb01280.x

Table 1

Overview of reliability, mean differences of the IRAP, VT and CRT as a function of sexual orientation, and criterion validity (AUC).

	α	Heterosexual		Gay		Group Differences		AUC
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	$t_{(120)}$	<i>d</i>	
IRAP ^a	.78	0.17	0.32	-0.41	0.41	-8.39***	-1.58	.90***
VT	.90	0.96	0.49	-0.78	0.48	-17.79***	-3.59	.99***
CRT ^b	.03	-0.19	0.25	-0.22	0.22	-0.73	-0.12	.55

Note. ^a $df = 116$, ^b $df = 117$

*** $p < .001$

Table 2

Intercorrelations of difference scores of all indirect latency-based measures, participant sexual orientation, and age (below main diagonal partial correlations corrected for age).

	1	2	3	4	5
1. Age		.23*	-.15	-.15	.03
2. Sexual Orientation	-		-.62***	-.85***	-.07
3. IRAP ^d ^a	-	-.61***		.55***	.12
4. VT ^d	-	-.86***	-.54***		.09
5. CRT ^d ^b	-	-.10	.14	.11	

Note. $N = 122$, ^a $N = 118$, ^b $N = 119$, ^c $N = 121$

* $p < .01$, *** $p < .001$

Figure 1. Mean d-IRAP scores as a function of trial combinations and sexual orientation. D-IRAP Scores are calculated by subtracting mean response latencies in heterosexual-congruent trial blocks from mean latencies of gay-congruent blocks. Error bars represent standard errors of means.

